Simplifying Explanations in Bayesian Belief Networks

Luis M. de Campos[1], José A. Gámez[2], and Serafín Moral[1]

[1] Dpto. de Ciencias de la Computación e I.A. Universidad de Granada.
18071 - Granada - SPAIN
{lci,smc}@decsai.ugr.es
[2] Dpto. de Informática. Universidad de Castilla-La Mancha.
02071 - Albacete - SPAIN
jgamez@info-ab.uclm.es

Abstract. Abductive inference in Bayesian belief networks is intended as the process of generating the K most probable configurations given an observed evidence. These configurations are called *explanations* and in most of the approaches found in the literature, all the explanations have the same number of literals. In this paper we study how to simplify the explanations in such a way that the resulting configurations are still accounting for the observed facts.

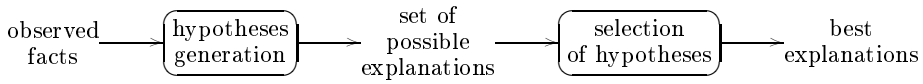
1 Introduction

Abduction [12] is defined as the process of generating a plausible explanation for a given set of observations or facts. This kind of reasoning can be represented by the following inference rule:

$$\frac{\psi \rightarrow \omega, \omega}{\psi},$$

i.e., if we observe ω and we have the rule $\psi \rightarrow \omega$, then we can infer that ψ is a *plausible* hypothesis (or explanation) for the occurrence of ω.

In general, there are several possible abductive hypotheses and it is necessary to choose among them. Therefore, we can divide the abductive task in two phases:



In order to select the best explanations from the generated set, two kinds of criteria are used: (1) metric based criteria (probability, weight, ...) and (2) simplicity criteria (the preferred explanation is the simplest available hypothesis). Usually *simplicity* is interpreted as logical simplicity, which means that those hypothesis with less different predicates are selected.

We think that the important role played by *simplicity* in the framework of abductive inference in logic has not been taken into account in the framework of

abductive inference in Bayesian belief networks. In this paper simplicity criteria that can be used in the context of Bayesian belief networks are studied.

The most similar work that we can find in the literature is due to Shimony [16, 17]. In this work, the explanation is a truth assignment to the *relevant* nodes in the network. That is, although Shimony works with partial abduction, he does not start with an explanation set, on the contrary he tries to identify the relevant nodes directly. The relevant nodes include the evidence nodes in the network and only ancestors of evidence nodes can be relevant (see [16, 17] for details). In our opinion the advantage of the Shimony's method is that it does not need two steps in the inference process, i.e., the method directly generates simplified explanations. However, as Chajewska and Halpern point out in [1] the explanations obtained by this method are not necessarily as concise as one could expect. In fact, it is not difficult to see that for each evidence node X_i, the explanation must include an assignment to all the nodes in at least one path from X_i to the root, since for each relevant node, at least one of its parents must be relevant.

The paper is organized as follows: In the second section we introduce abductive inference in Bayesian belief networks. In the third section, we propose two kinds of simplification criteria. In the fourth section computation issues are studied. Finally, in the fifth section, we consider the conclusions.

2 Abductive Inference and Bayesian Belief Networks

A *Bayesian belief network* (Pearl [11], Jensen [6]) is a directed acyclic graph (DAG) where each node represents a random variable, and the topology of the graph shows the (in)dependence relations among the variables. The quantitative part of the model is given by a probability distribution for each node conditioned to its parents. If $X_{\mathcal{U}} = \{X_1, \ldots, X_n\}$ is the set of variables in the network, then the join probability can be calculated as:

$$p(X_{\mathcal{U}}) = \prod_{X_i \in X_{\mathcal{U}}} p(X_i | pa(X_i)),$$

where $pa(X_i)$ contains the parents of X_i.

Given an evidence $X_O = x_O$, propagation algorithms allow to calculate $p(X_i|x_O)$ for every $X_i \in X_{\mathcal{U}} \setminus X_O$. The calculations are carried out in a secondary structure (obtained from the original BBN) called a *junction tree*, where the evidence x_O has been inserted. The method is based on the use of two operations: *marginalization* (addition) and *combination* (multiplication), see [6, 14] for details.

In the context of BBNs abductive inference corresponds to finding the maximum a posteriori probability state of the network, given the instantiated variables (the evidence). If X_O is the set of observed variables and X_U is the set of unobserved variables, we want to obtain the configuration x_U^* of X_U such that:

$$x_U^* = \arg\max_{x_U} P(x_U|x_O), \tag{1}$$

where $X_O = x_O$ is the observed evidence. Usually, x_U^* is known as the *most probable explanation* (MPE), and in general we are interested in the K most probable explanations (K MPEs).

Sometimes we are interested in obtaining the K MPEs only for a subset of the network's variables called *explanation set* [9]. This problem is known as *Partial Abductive Inference* and we think that in practical applications is more interesting than the classical abductive inference problem, because we can select as the explanation set the variables representing diseases in a medical diagnosis problem, the variables representing critical components (starter, battery, alternator, ...) in a car diagnosis problem, etc ....

Now, if we denote by $X_E \subset X_U$ the explanation set, then we want to obtain the configuration x_E^* of X_E such that:

$$x_E^* = \arg\max_{x_E} P(x_E|x_O) = \arg\max_{x_E} \sum_{x_R} P(x_E, x_R|x_O), \tag{2}$$

where $X_R = X_U \setminus X_E$. In general, x_E^* is not equal to project the configuration x_U^* over the variables of X_E. Therefore, we need to obtain x_E^* directly (eq. 2).

The MPE $x_{\mathcal{U}}^*$ can be found using the *probabilities propagation* method replacing addition by maximum in the marginalization operator [2]. To obtain the K MPEs more complex methods must be used ([10, 13]). The process to find the MPE x_E^* is more complex because not all junction trees obtained from the original BBN are valid. In fact, because addition and maximum have to be used simultaneously, the variables of X_E must form a subtree of the complete junction tree for $X_{\mathcal{U}}$. Gámez [4] has shown that in this case the size of the junction tree grows in relation to the size of the junction tree obtained without restrictions, and so the propagation algorithm will be less efficient. In [3] an approximate method based on genetic algorithms is proposed.

3 Simplicity Criteria

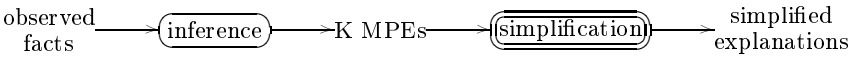
As we have seen in the previous section, when abductive inference is carried out in BBNs a *metric* is used to select the best explanations, concretely the explanations are ranked by their *a posteriori probability*. However, no simplicity criterion is applied. An immediate consequence is that all explanations have the same number of literals ($|X_E|$[1]). Thus, if we have observed *car does not start* and our explanation set contains critical components of the car, we can get

battery=dead, alternator=ok, starter=ok, engine cranks=ok, ...

as a MPE. However, the explanation *battery=dead* could be enough to account for the observation and is simpler than the previous one. Therefore, our goal is to simplify the explanation obtained applying the methods cited in the previous

[1] In the rest of the paper we are not going to distinguish between abductive inference and partial abductive inference, simply we suppose $X_E = X_U$ for the case of (complete) abductive inference

section, removing from the explanation those literals that are not important given the evidence. The process can be represented as follows



and can be stated in a more formal way as:

> Let $expl(x_O) = \{x_E^1, x_E^2, \ldots, x_E^K\}$ be the K MPEs obtained for the evidence $X_O = x_O$. Then, for all $x_E \in expl(x_O)$ we are looking for a sub-configuration x'_E ($X'_E \subset X_E$), such that x'_E is still accounting for the evidence.

We are going to note by $x'_E \subset x_E$, that x'_E is obtained from x_E by removing one or more literals; and by $|x_E|$ the number of literals in x_E.

The next two sections are devoted to introduce two kinds of criteria that can be used to decide when $x'_E \subset x_E \in expl(x_O)$ is still an explanation for x_O.

3.1 Independence Based Criteria

Suppose that we can divide our initial explanation x_E in two parts, x_D and x_I ($X_D \cup X_I = X_E$), such that if we know x_D then adding x_I to our knowledge does not modify our belief on the presence of the evidence (x_O). Then, we can say that x_I is irrelevant for the evidence and remove its literals from the explanation. From the probabilistic point of view, we can express this idea as follows:

$$P(x_O|x_D) = P(x_O|x_D, x_I). \tag{3}$$

Following this idea we can give the next definition of simplification:

Definition 1. *(I-simplification)*
We say that $x'_E \subset x_E$ is an Independence based simplification *$(I-simplification)$ of $x_E \in expl(x_O)$ if and only if $P(x_O|x'_E) = P(x_O|x_E)$.*

In order to relax the previous definition the term *equal* can be replaced by *almost equal*. Thus, we have the following definition:

Definition 2. *(I∼simplification)*
We say that $x'_E \subset x_E$ is an Independence based simplification *(I∼simplification) of $x_E \in expl(x_O)$ if and only if $P(x_O|x'_E) \simeq P(x_O|x_E)$.*

The admissible difference between $P(x_O|x_E)$ and $P(x_O|x'_E)$ can be fixed by a threshold, i.e., we can take $\epsilon \in [0.01, 0.05]$ and use the following equation:

$$(1-\epsilon)P(x_O|x_E) \leq P(x_O|x'_E) \leq (1+\epsilon)P(x_O|x_E) \tag{4}$$

Before going on with our study, it is convenient to say that similar definitions to the previously formulated can be found in works devoted to *sensitivity analysis in Bayesian networks* (see [7, 18]). However, our idea should be interpreted in the opposite sense, because sensitivity analysis studies how sensitive is the conclusion

(hypothesis) with respect to the set of observations, analyzing which items of evidence are *in favour of/against/irrelevant for* the conclusion. In our case, the evidence is the only thing we consider previously fixed, and we want to analyze which subsets of the hypothesis are still an explanation for the given evidence.

Taking up again our definition of I∼simplification, we can see that some explanations can not be simplified, but others can have more than one simplification. In the last case, we want to obtain the *best* possible simplification, i.e., the simplification with the smallest number of literals. The following definition formalize this idea and uses probability to break ties:

Definition 3. *(Best I∼simplification)*
We say that $x'_E \subset x_E$ *is the* best independence based simplification *(I∼simplification) of* $x_E \in expl(x_O)$ *if and only if the following conditions hold:*

1. x'_E *is an I∼simplification of* x_E.
2. $\nexists\ x''_E \subset x_E$, *such that is an I∼simplification of* x_E *and* $|x''_E| < |x'_E|$.
3. *If* x''_E *is an I∼simplification of* x_E *and* $|x''_E| = |x'_E|$ *then the following expression is true:* $abs(P(x_O|x''_E) - P(x_O|x_E)) \geq abs(P(x_O|x'_E) - P(x_O|x_E))$.

In section 4 we will talk about computational aspects of this criterion, but now we are going to give another independence based criterion.

Now, our idea is the following. Let x'_E be an I∼simplification of x_E, then we have removed the literals in $X_E \setminus X'_E$ because they were (almost) irrelevant for the observed evidence. But, what happen when $P(x_O|x_E)$ is not (almost) equals to $P(x_O|x'_E)$?. We can distinguish two cases:

1. $P(x_O|x'_E) < P(x_O|x_E)$. This can be interpreted as the sub-configuration x'_E account for x_O in smaller degree than the original explanation.
2. $P(x_O|x'_E) > P(x_O|x_E)$. This can be interpreted as the sub-configuration x'_E account for x_O in greater degree than the original explanation.

Therefore, in the second case we think that x'_E would be accepted as a simplification of x_E, because it implies the observed evidence at least at the same degree as x_E and it has a smaller number of literals. This idea can by formalized by the following definition:

Definition 4. *(R∼simplification)*
We say that $x'_E \subset x_E$ *is an* Relevance based simplification *(R∼simplification) of* $x_E \in expl(x_O)$ *if and only if* $P(x_O|x'_E) >\approx P(x_O|x_E)$.

We have called this criterion, *relevance* based simplification because the removed literals are not irrelevant to the evidence, against what happen when using I∼simplification. It is clear that the *best* R∼simplification can be defined similarly to definition 3, breaking ties in favour of the sub-configuration x'_E with greatest probability $P(x_O|x'_E)$.

3.2 Normality Based Criteria

One of the main applications of abduction is in the field of diagnostic reasoning. In this field, we can simplify the explanation, giving as result only those variables which are not OK, that is, we can simplify the explanation by omitting those variables which are taking their *usual* state.

For some variables the *usual* state is specified in the model, for example, in medical diagnosis the usual state for variables representing diseases is *absent*, and in circuit diagnosis the usual state for logical gates and components is *no fault*. However, we can formulate us the next two questions:

1. what's happen with those variables whose usual state is unknown?, and
2. does an usual state exist for each variable of the model?

We think that the answer to the second question is NOT. For example, if we consider the variable *sex*, both states *male* and *female* have the same likelihood. In our opinion, variables without usual state do not have to be removed from the initial explanation. For the first question, we are going to give criteria to estimate the most usual value of a variable (if exists). Before that, the next two definitions state how to obtain a normality based simplification:

Definition 5. *(Normalization Mask)*
Let $X_E = \{E_1, \ldots, E_n\}$ be the explanation set. Let $\Omega_{E_i} = \{e_i^1, \ldots, e_i^{k_i}\}$ be the set of possibles states for E_i. Then, the normalization mask*, noted by $Mask(X_E)$, is an array of n elements, such that:*

$$Mask(X_E)[i] = \begin{cases} e_i^j & \text{if } e_i^j \text{ is the usual state for } E_i \\ \# & \text{if no usual state exists for } E_i \end{cases}$$

Definition 6. *(Normality Based Simplification)*
Given an explanation $x_E \in expl(x_O)$, its normality based simplification $simp(x_E)$ is the vector obtained from X_E by deleting the E_i component, $x_E^{\downarrow E_i}$, if it coincides with the usual state of E_i: $Mask(X_E)[i]$, for every $E_i \in X_E$.

Example 1. Let $X_E = \{\underline{C}old, \underline{F}lu, \underline{S}moker, \underline{B}ronchitis, \underline{T}uberculosis\}$ be the explanation set. Let $Mask(X_E) = [\overline{c}, \overline{f}, \#, \overline{b}, \overline{t}]$ be the normalization mask (all the variables can take two states $\{a, \overline{a}\}$, representing presence or absence respectively). Then, the simplification of $x_E = (\overline{c}, f, s, b, \overline{t})$ is (f, s, b).

Now our goal is to obtain the normalization mask. We are going to approach the problem from two distinct points of view: local and global.

Local Estimation of the Normalization Mask. From *local* we are understanding that the *usual* state of a variable is determined without taking into account the rest of variables in the explanation set. Thus, our definition of *usual* state is based in the *a priori* marginal probability of each variable.

Definition 7. *(Usual State)*
Let X_i be a variable taking values in $\Omega_{X_i} = \{x_i^1, \ldots, x_i^{k_i}\}$. We say that x_i^j is the usual state *of X_i, if and only if the following conditions hold:*

1. $\forall l \neq j, P(X_i = x_i^j) \geq 2P(X_i = x_i^l)$, *and*
2. $P(X_i = x_i^j) \geq \sum_{l \neq j} P(X_i = x_i^l)$.

If the previous conditions do not hold for any state of X_i *then we say that this variable has not an* usual state.

The first condition requires a significant difference between the probability of the usual state and the rest. The second condition is introduced to avoid cases as the following: Let us consider $\Omega_{X_i} = \{x_i^1, \ldots, x_i^8\}$, $P(x_i^1) = 0.2$ and $P(x_i^j) = 0.1$ for $j = 2, \ldots, 8$; then x_i^1 satisfies the condition (1), but we think that this state should not be considered as the usual state for X_i because the probability of X_i taking any other state is four times the probability of X_i taking the state x_i^1.

Global Estimation of the Normalization Mask. From *global* we are understanding that the *usual* state of a variable is determined taking into account the rest of variables in the explanation set. Thus, our definition of *usual* state is based in the *a priori* joint probability. That is, to obtain the normalization mask we use the configurations of X_E with high value of $P(X_E|\emptyset)$. Now, the question is how many configurations must be used?.

The answer to the previous question is not easy, but we think that a threshold can be used. For example, if x_E^1 is the configuration with highest value of $P(X_E|\emptyset)$, then we can use the configurations whose probability is between $P(x_E^1)$ and $\alpha \cdot P(x_E^1)$, with $\alpha \geq 0.5$. Using this criterion the normalization mask can be obtained as follows:

$$Mask(X_E)[i] = \begin{cases} e_i^j & \text{if } x_E^{\downarrow E_i} = e_i^j \text{ for all the considered configurations} \\ \# & \text{otherwise} \end{cases}$$

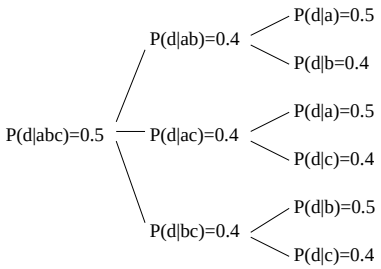
4 Computation

In normality based criteria the key point is the calculation of the normalization mask. In local estimation, a priori *marginal* probabilities must be calculated ($P(X_i)$ for all $X_i \in X_E$), it is clear that this task can be achieved by probabilities propagation methods (for example HUGIN [8]). In global estimation we need to calculate joint probabilities for the variables in the explanation set, so this is an abductive inference problem without observed evidence and the task can be achieved by using the methods cited in Section 2.

In independence based criteria the computation is much more complicated. First, the search can not be done by removing a literal at each step, because the independence based criteria does not have a monotonicity property, as we can see in the following example.

Example 2. Let us to consider the network given by the graph $G = (\{A, B, C, D\}, \{A \rightarrow D, B \rightarrow D, C \rightarrow D\})$ and the conditional probabilities shown in the following table (all the variables can take two states).

$P(\mathtt{a}) = 0.5$	$P(\mathtt{d}\|\mathtt{abc}) = 0.5$	$P(\overline{\mathtt{d}}\|\mathtt{abc}) = 0.5$
$P(\overline{\mathtt{a}}) = 0.5$	$P(\mathtt{d}\|\mathtt{ab}\overline{\mathtt{c}}) = 0.3$	$P(\overline{\mathtt{d}}\|\mathtt{ab}\overline{\mathtt{c}}) = 0.7$
	$P(\mathtt{d}\|\mathtt{a}\overline{\mathtt{b}}\mathtt{c}) = 0.3$	$P(\overline{\mathtt{d}}\|\mathtt{a}\overline{\mathtt{b}}\mathtt{c}) = 0.7$
$P(\mathtt{b}) = 0.5$	$P(\mathtt{d}\|\mathtt{a}\overline{\mathtt{b}}\overline{\mathtt{c}}) = 0.9$	$P(\overline{\mathtt{d}}\|\mathtt{a}\overline{\mathtt{b}}\overline{\mathtt{c}}) = 0.1$
$P(\overline{\mathtt{b}}) = 0.5$	$P(\mathtt{d}\|\overline{\mathtt{a}}\mathtt{bc}) = 0.3$	$P(\overline{\mathtt{d}}\|\overline{\mathtt{a}}\mathtt{bc}) = 0.7$
	$P(\mathtt{d}\|\overline{\mathtt{a}}\mathtt{b}\overline{\mathtt{c}}) = 0.5$	$P(\overline{\mathtt{d}}\|\overline{\mathtt{a}}\mathtt{b}\overline{\mathtt{c}}) = 0.5$
$P(\mathtt{c}) = 0.5$	$P(\mathtt{d}\|\overline{\mathtt{a}}\overline{\mathtt{b}}\mathtt{c}) = 0.5$	$P(\overline{\mathtt{d}}\|\overline{\mathtt{a}}\overline{\mathtt{b}}\mathtt{c}) = 0.5$
$P(\overline{\mathtt{c}}) = 0.5$	$P(\mathtt{d}\|\overline{\mathtt{a}}\mathtt{bc}) = 0.5$	$P(\overline{\mathtt{d}}\|\overline{\mathtt{a}}\mathtt{bc}) = 0.5$



Let $D = d$ be the evidence and $X_E = \{A, B, C\}$ be the explanation set. Then, abc is the second most probable explanation with $P(\mathtt{abc}|\mathtt{d}) = 0.133$. In the tree obtained when all the sub-configurations of abc are considered we can see that ab is not an independence (or relevance) based simplification for abc, while its sub-configuration a is a valid simplification.

Therefore, there is not monotonicity in the space of sub-configurations and the complete search space should be explored in order to find the *best* independence (or relevance) based simplification. So, we need to calculate $P(x_O|x'_E)$ for each $x'_E \subset x_E$, and given that the number of sub-configurations grows exponentially with the number of variables in the explanation set, the process could be intractable.

In sensitivity analysis [7] the same problem exists, and the solution adopted by Jensen et al. was to use a modified scheme of inference in junction trees called *cautious propagation* (Jensen, [5]). Cautious propagation is a modification of HUGIN propagation into a Shafer-Shenoy-like architecture [15]; it is less efficient than HUGIN, but combined with cautious entering of evidence it provides access to $P(e'|h)$ for a great number of subsets e'. As $P(x_O|x'_E) = \frac{P(x'_E|x_O)P(x_O)}{P(x'_E)}$ we can use this method but doing two cautious propagations, one without entering evidence -to obtain $P(x'_E)$- and other entering evidence -to obtain $P(x'_E|x_O)$-. The solution obtained by this method can be interpreted as approximate, because the search space is not completely explored.

In [4] an alternative approximate method is proposed. It is based on an incremental heuristic search. We start with the complete explanation and we keep on removing a literal in each step. The procedure stops when it is not possible to obtain a simplification by deleting a single literal. In this way the number of evaluated sub-configurations is reduced from $2^{|X_E|}$ to $\frac{|X_E|^2+|X_E|}{2}$. Furthermore, this is an upper bound because the explanation is not always simplified to an unique literal. As $P(x_O|x'_E) = \frac{P(x'_E,x_O)}{P(x'_E)}$ two ascendent propagations (only the first step of HUGIN is carried out) are necessary in order to calculate this value (with and without entered evidence). However, when the initial explanation x_E has been evaluated, the messages sent in propagation are stored in order to avoid repetitions of computations when their sub-configurations are being evaluated.

Finally, we think that in most cases a first independence (or relevance) based simplification can be carried out by studying the graph, i.e., without probabilistic

propagations. The idea [4] is to find the minimal subset X_D of X_E such that the following independence sentence is true:

$$I(X_O|X_D|X_I),$$

with $X_I = X_E \setminus X_D$. In this case, it is clear that $I(x_O|x_D|x_I)$ holds for each sub-configuration x_O, x_D, x_I of X_O, X_D and X_I respectively. In [4] an algorithm is proposed in which X_D can be found using *d-separation* [11].

5 Concluding Remarks

In this paper several simplification criteria have been proposed. Normality based criteria have the advantage of being much more easily calculated than independence based criteria. On the other hand, we think that independence based criteria have a more consistent semantic. In the experiments that we have carried out with real and artificial BBNs the number of literals in the simplified explanations is between the 15% and the 35% of the literals in the initial explanation, when using independence based criteria, and between the 22% and the 65% when using normality based criteria.

In order to show a real example of our simplification criteria, we can consider the following case (taken from one of the experiments carried out over the network *car-starts*[2]. The root nodes were selected as the explanation set, and *Lights=work, Starts=No* as the evidence. Then, the obtained MPE is:

$$Alternator = Ok, FanBelt = Ok, Leak = NoLeak, BatteryAge = New,$$
$$GasInTank = NoGas, Starter = Faulted, Leak2 = False, FuelPump = Ok,$$
$$Distributor = OkandSparkPlugs = Ok.$$

And the simplified explanation are[3]:

Normality: $GasInTank = NoGasandStarter = Faulted$

Independence: $BatteryAge = NewandStarter = Faulted.$

The complexity of computation in independence based criteria makes necessary to use some kind of approximate methods. In this case, an adaptation of Jensen's cautious propagation can be used. In the future, we plan to devote more effort to the computational issues.

We have detected that if the explanation set is not carefully selected then independence based criteria can yield immediate explanations, i.e., if you observe that the car does not start, the explanation would be *no gas in tank*, but not the reason because there is no gas in tank. To avoid this problem we plan to develop an iterative method in order to find more specific explanations.

Given the complexity of abductive reasoning in BBNs we plan to investigate in the development of approximate methods that directly yield simplified explanations, i.e., the process only will involve one step and not two as in the present work.

---

[2] This network is included in the package JavaBayes `www.cs.cmu.edu/ javabayes`

[3] In this case both independence criteria (I,R) yield the same simplification. This fact also occurs with the two types of mask estimation

6 Acknowledgments

This work has been supported by the Spanish Comisión Interministerial de Ciencia y Tecnología (CICYT) under Project TIC97-1135-CO4-01.

References

[1] U. Chajewska and J. Y. Halpern. Defining explanation in probabilistic systems. In *Proceedings of the Thirteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI–97)*, pages 62–71, San Francisco, CA, 1997. Morgan Kaufmann.

[2] A.P. Dawid. Applications of a general propagation algorithm for probabilistic expert systems. *Statistics and Computing*, 2:25–36, 1992.

[3] L.M. de Campos, J.A. Gámez, and S. Moral. Partial Abductive Inference in Bayesian Belief Networks using a Genetic Algorithm. Patter Recognition in Practice VI. Vlieland, The Netherlands., 1999.

[4] J.A. Gámez. *Inferencia abductiva en redes causales.* PhD thesis, Departamento de Ciencias de la Computación e I.A. Universidad de Granada, 1998.

[5] Finn V. Jensen. Cautious propagation in Bayesian networks. In *Proceedings of the Eleventh Annual Conference on Uncertainty in Artificial Intelligence (UAI–95)*, pages 323–328, Montreal, Quebec, Canada, 1995.

[6] F.V. Jensen. *An introduction to Bayesian Networks.* UCL Press, 1996.

[7] F.V. Jensen, S.H. Aldenryd, and K.B. Jensen. Sensitivity analysis in Bayesian networks. In *Symbolic and Quantitative Approaches to Reasonning and Uncertainty*, pages 243–250. Springer Verlag LNAI 946, 1995.

[8] F.V. Jensen, S.L. Lauritzen, and K.G. Olesen. Bayesian updating in causal probabilistic networks by local computation. *Computational Statistics Quarterly*, 4:269–282, 1990.

[9] R. E. Neapolitan. *Probabilistic Reasoning in Expert Systems. Theory and Algorithms.* Wiley Interscience, New York, 1990.

[10] D. Nilsson. An efficient algorithm for finding the m most probable configurations in Bayesian networks. *Statistics and Computing*, 2, 1998.

[11] J. Pearl. *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufmann, 1988.

[12] H.E. Pople. On the mechanization of abductive logic. In *Proceedings of the 3rd International Joint Conference on Artificial Intelligence*, 1973.

[13] B. Seroussi and J.L. Goldmard. An algorithm directly finding the k most probable configurations in Bayesian networks. *International Journal of Approximate Reasoning*, 11:205–233, 1994.

[14] G. Shafer. *Probabilistic Expert Systems.* Society for Industrial and Applied Mathematics (SIAM), 1996.

[15] P.P. Shenoy and G.R. Shafer. Axioms for probability and belief-function propagation. In R.D. Shachter, T.S. Levitt, L.N. Kanal, and J.F. Lemmer, editors, *Uncertainty in Artificial Intelligence, 4.*, pages 169–198. Elsevier Science Publishers B.V. (North-Holland), 1990.

[16] S.E. Shimony. The role of relevance in explanation I: Irrelevance as statistical independence. *International Journal of Approximate Reasoning*, 8:281–324, 1993.

[17] S.E. Shimony. The role of relevance in explanation II: Disjunctive assignments and approximate independence. *International Journal of Approximate Reasoning*, 13:27–60, 1995.

[18] H.J. Suermondt. Explanation of probabilistic inference in Bayesian belief networks. Technical Report KSL-91-39, Stanford University, 1991.